DAILY BEACH WATER QUALITY FORECAST: 3D DETERMINISTIC MODEL VS STATISTICAL MODEL

K.W. CHOI

Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China, dkwchoi@ust.hk

JOSEPH H.W. LEE

Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China, jhwlee@ust.hk

ABSTRACT

Escherichia coli (*E. coli*) concentration is adopted as the main indicator of beach water quality in Hong Kong due to its high correlation with swimming associated illnesses. As part of the WATERMAN system - coastal water quality forecast and management system for Hong Kong, both 3D deterministic hydrodynamic and multiple linear regression (MLR) models have been developed to provide daily water quality forecasting for eight marine bathing beaches in Tsuen Wan, which are only about 8 km from the Harbour Area Treatment Scheme (HATS) outfall discharging 2.5 million m^3/d of chemically enhanced primary treatment (CEPT) effluent. The forecast performance of the two models on the compliance of beach water quality objective is studied for a three-year period (2016-2018). While the 3D model is process-based and the MLR model is data-driven, both models have comparable performance with an overall accuracy of about 70-80%. The model forecasts of *E. coli* concentrations are significantly correlated with the sampling data obtained from the regular monitoring programme. In general, the MLR models have slightly higher overall accuracy and better correlation with the observation, but by its very nature can predict the *E. coli* level only around the standard sampling time on daily basis (around 11 am). On the other hand, the 3D model can capture the diurnal variation of beach water quality and provides better prediction for the observed water quality exceedances.

Keywords: beach water quality, hydrodynamic model, statistical model, real-time forecast

1. INTRODUCTION

Over the past two decades a great deal of research has demonstrated the dynamic variability of beach water quality as a function of environmental factors and pollution sources. There is also a growing awareness on the advantages of the use of predictive models to supplement routine monitoring. In 2008, a pioneering water quality forecast and management system for Hong Kong (entitled Project "WATERMAN") was initiated. The objective of the project is to develop an innovative internet and GIS-based environmental knowledge base for Hong Kong. One of the three core modules of the WATERMAN system is a beach water quality forecast system which have been providing regular (daily) water quality forecasts for Hong Kong bathing beaches (Thoe *et al.* 2012, Chan *et al.* 2013, Thoe and Lee 2014).

In Hong Kong, epidemiological studies have shown that *Escherichia coli* (*E. coli*) concentration is a good indicator of faecal pollution; *E.coli* level showed a strong relationship with the incidence rate of swimming-associated illnesses such as skin and gastro-intestinal illnesses (Cheung *et al.* 1990). A beach with *E. coli* level exceeding 180 counts/100 mL will be considered one with "poor" water quality.

The Hong Kong Harbour Area Treatment Scheme (HATS) is a major environmental infrastructure that aims to improve the water quality of Victoria Harbour. Wastewater from the urban areas surrounding the harbour is collected through a deep tunnel system and conveyed to the Stonecutters Island Sewage Treatment Work (SCISTW) for centralized chemically enhanced primary treatment (CEPT). The treated effluent from SCISTW (2.5 million m^3/d) is discharged to the waters southwest of Stonecutters Island from a 1.2 km long submerged diffuser with rosette type risers. A group of eight beaches in the Tsuen Wan district is located just about 8 km to the northwest of this sewage outfall (Fig. 1), to the west of Victoria Harbour and near the Tsing Yi Island. Seven of the beaches, namely Anglers' (ANG), Gemini (GEM), Hoi Mei Wan (HOM), Casam (CAS), Lido (LIDO), Ting Kau (TK) and Approach (APP) beaches are located on the coastline to the west of Tsuen Wan Town along the Rambler Channel. In contrast, Ma Wan Tung Wan (TWM) Beach is located on Ma Wan Island to the west of Ma Wan Channel. It is expected that water quality in these beaches would be significantly influenced by this single major pollution source. Since March 2011 the treated CEPT effluent is also disinfected by chlorination. The beach water quality has improved greatly since then; the changes in bacterial levels and recent data offer a good opportunity to study the performance of predictive models.

2. MATHEMATICAL MODELS

The beach water quality module of the WATERMAN system has two modelling components:

(a) Statistical daily water quality prediction system: Based on regular monitoring data of the Hong Kong Environmental Protection Department (EPD), Multiple Linear Regression (MLR) models have been developed individually for each of the 43 bathing beaches in Hong Kong. The models predict the *E. coli* level for each beach around the standard sampling time (around 11 am) on daily basis (Thoe *et al.* 2012). The daily beach water quality forecasts have been systematically validated against field data (Thoe and Lee 2014). As field observations show that the *E. coli* concentration on a beach is typically log-normally distributed, the dependent variable to be predicted is taken to be the *E. coli* concentration in natural logarithm (lnEC) (Thoe *et al.* 2012). It is assumed that lnEC is a linear combination of selected hydro-environmental variables:

$$Y = b_0 + \sum_{j=1}^{M} b_j x_j \tag{1}$$

where Y is the predicted value of lnEC, x_j is value of jth variable, and b_j is the regression coefficient of the jth variable, and b_0 is the constant term. The variables employed in the WATERMAN MLR model include rainfall, solar radiation, predicted tide level, prevailing onshore wind speed, salinity, water temperature and previous *E. coli* concentration.

Based on EPD monitoring data (at intervals of 3-14 days during the bathing season) over a 5 year-period after the introduction of disinfection (2011-2015), individual MLR models are developed for each of the eight Tsuen Wan beaches.

(b) 3D hydrodynamic and water quality model: The model is based on a unique dynamic coupling of a Lagrangian near field buoyant jet model JETLAG (Lee and Chu 2003) and the Environmental Fluid Dynamics Code (EFDC) far field shallow water circulation model. With the use of the Distributed Entrainment Sink Approach (DESA), the near field jet model predictions for the initial jet mixing zone can be incorporated seamlessly into the far field model at *grid cell level* (Choi and Lee 2007; Lee *et al* 2013). The *E. coli* decay formulation in the model is derived from laboratory experiments and field studies for Hong Kong coastal water (Chan *et al.* 2015). The model prediction of beach water quality has been extensively validated against field data both before and after disinfection of the HATS effluent. Unlike the statistical approach, the model embodies the bacterial loadings from all pollution sources and solves for the flow and *E.coli* concentration for the entire Hong Kong waters. In particular, as shown in Fig. 2, the model can give real time (e.g. hourly) predictions of beach water quality at the Tsuen Wan beaches (Chan *et al.* 2013).

3. RESULTS AND DISCUSSION

3.1 Metrics for assessing model forecast performance

A key performance indicator of a beach water quality model is the accuracy of the predicted compliances/exceedances of the water quality criterion as defined by the threshold *E.coli* concentration (180 counts/100mL). Based on long term field observation and modelling experience, the following assessment criteria are considered: correlation coefficient, overall accuracy (the percentage of correct forecast of compliance/exceedance of the threshold level) .and sensitivity (the percentage of observed exceedances that are actually predicted).

3.2 The daily forecast performance for the three-year period (2016 - 2018)

To assess the performance of the two models, the model results are compared against the EPD regular beach monitoring data for the recent three years 2016 – 2018 that have not been used in the training or calibration of the MLR model. The total number of observations in the three-year period for each beach is about 115.

As shown in Table 1, it can be seen that the overall accuracy for all the beaches are around 70% - 80% and the correlation between the model and observed results of the MLR models is very good, in the range of 0.34 to

0.52. It should be noted that the correlation is superior to the existing beach management approach based on the geometric mean of the 5 previous measurements; with no forecasting ability the correlation coefficient (not shown) is around -0.2 to 0.2 for the same beaches. Except for the clean beaches TWM and ANG, the performance of the 3D model is comparable to that of MLR. The sensitivities (the percentage of correct prediction of observed exceedance) of both models are less satisfactory, in particular, for the cleaner beaches (such as TWM and LIDO) with less number of observed exceedances. The 3D model has generally better sensitivity than the MLR model. The above indicate that the MLR models capture the general trends and the averaged behaviour better, whereas the 3D model represents the extreme events better.

In particular, TWM is a relatively "clean" beach with annual geometric mean *E. coli* concentration mostly below 40 counts/100mL between 2010 and 2018. The water quality on such a beach is very much governed by short-term variations in unknown local non-point pollution sources (e.g. due to rainfall) that are not readily accounted for in the deterministic 3D model – in general the predicted *E. coli* concentrations are lower than observed. This is supported by the significantly higher correlation for the cases with negligible prior rainfall (Fig. 3).

Figure 1. Location of the Tsuen Wan Beaches and HATS sewage outfall.

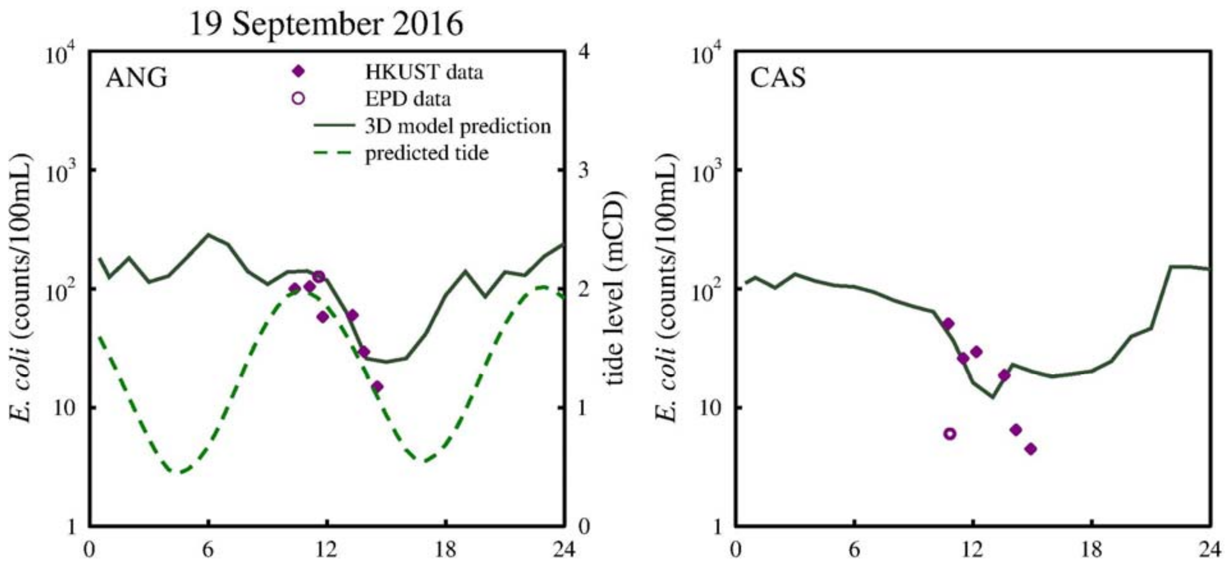

Figure 2. Predicted diurnal variations of *E. coli* at Anglers' and Casam beach on 19 Sep 2016.

Fig. 4 shows the measured and predicted *E. coli* concentrations at Ting Kau Beach for June – October 2018 (bathing season). It can be seen that the forecast results of the two models show very similar time varying pattern. Both models capture the short-term *E. coli* variation in response to dynamic changes in hydro-meteorological and tidal conditions and environmental changes much better than the current EPD advisory (C_{lnEC5}) based on the geometric mean of the last 5 measurements.

Table 1. Forecast performance of the MLR and 3D models for Tsuen Wan beaches for 3-year period 2016 - 2018.

Beach	Correlation coefficient		Overall accuracy		Sensitivity (no. of correct forecast / no. of observed exceedances)	
	MLR model	3D model	MLR model	3D model	MLR model	3D model
TWM	0.352	0.112	81.2%	80.9%	2/20	0/20
ANG	0.462	0.157	78.6%	73.5%	10/32	4/32
GEM	0.384	0.265	74.8%	73.7%	2/29	3/29
HOM	0.454	0.445	75.9%	71.7%	4/30	11/30
CAS	0.370	0.387	79.1%	67.8%	5/26	8/26
LIDO	0.339	0.286	80.0%	66.7%	1/19	3/19
TK	0.519	0.484	75.6%	69.7%	16/37	24/37
APP	0.452	0.408	75.0%	70.1%	14/38	14/38

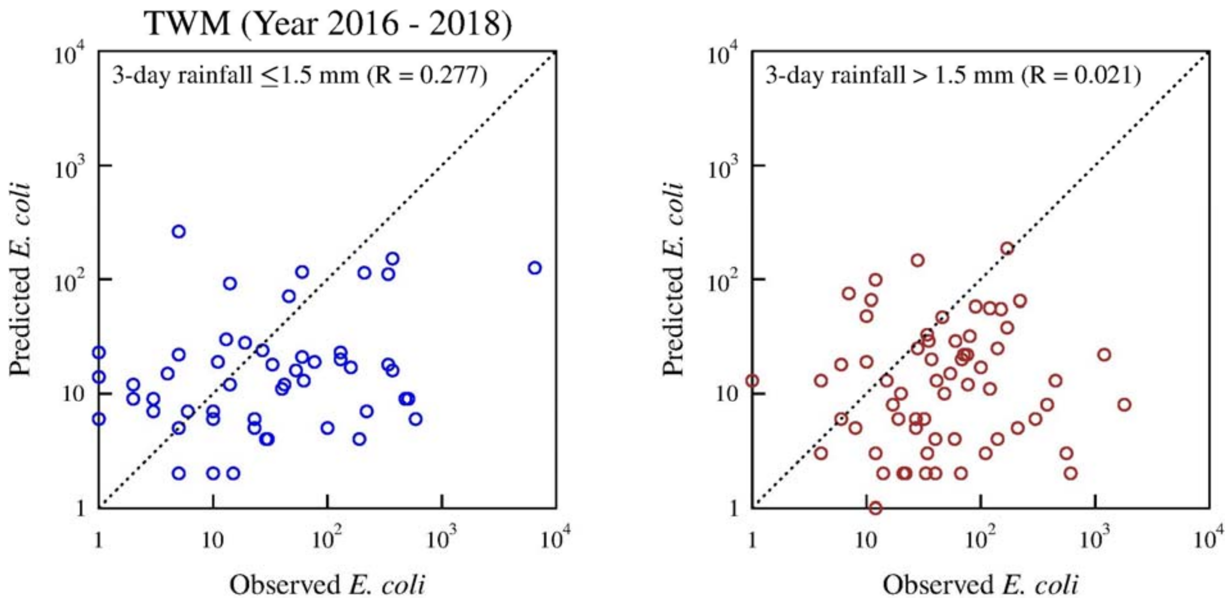

Figure 3. The 3D model predicted vs. observed *E. coli* for Ma Wan Tung Wan (TWM) beach for year 2016 – 2018.

4. CONCLUSIONS

The performance of statistical multiple linear regression (MLR) and 3D deterministic flow and water quality models for beach water quality forecasting is systematically studied against field data for the first time. The model forecasts of *E. coli* concentrations are significantly correlated with observations; the forecast performance is superior to traditional approaches that depend purely on past data. The MLR models have slightly higher overall accuracy and correlation with the observation, but can only predict the *E. coli* level at around the standard sampling time (11 am) on daily basis. On the other hand, the 3D model captures the diurnal variation of the beach water quality and provides better prediction for the observed water quality exceedances. The extensive model-data comparison indicates that the data-driven model captures the general trends better, whereas the 3D model represents the extreme events better.

ACKNOWLEDGMENTS

This work was supported by a research consultancy project ("Beach Water Quality Prediction System for HATS 2A Post Project Monitoring Study") commissioned by the Environmental Protection Department of the HKSAR Government. Part of the beach water quality data were provided by the Drainage Services Department of HKSAR Government.

REFERENCES

Chan, S.N., Thoe, W. and Lee, J.H.W. (2013). Real-time forecasting of Hong Kong beach water quality by 3D deterministic model, *Water Research*, 47(4): 1631-1647.

Chan, Y.M., Thoe, W. and Lee, J.H.W. (2015). Field and laboratory studies of *Escherichia coli* decay rate in subtropical coastal water. *Journal of Hydro-environment Research*, 9(1): 1-14.

Cheung, W.H.S., Chang, K.C.K., Hung, R.P.S. and Kleevens, J.W. (1990). Health effects of beach water pollution in Hong Kong, *Epidemiology and Infection*, 105: 139-162.

Choi, K.W. and Lee, J.H.W. (2007). Distributed entrainment sink approach for modelling mixing and transport in intermediate field, *Journal of Hydraulic Engineering*, *ASCE*, 133(7): 804-815.

Lee, J.H.W. and Chu, V.H. (2003). *Turbulent jets and plumes - a Lagrangian approach*, Springer/Kluwer.

Lee, J.H.W., Wong, K.T.M. and Choi, K.W. (2013). Forecasting and Management of Coastal Water Quality, *Handbook of Environmental Fluid Dynamics, Vol. 1*, edited by Fernando, HJS, CRC Press/Taylor & Francis Group, LLC: 75-90.

Thoe, W.. and Lee, J.H.W. (2014). Daily forecasting of Hong Kong beach water quality by Multiple Linear Regression (MLR) models, *Journal of Environmental Engineering*, *ASCE*, 140(3): 241-253.

Thoe, W., Wong, H.C., Choi, K.W. and Lee, J.H.W., (2012). Daily prediction of marine beach water quality in Hong Kong, *Journal of Hydro-environment Research*, 6: 164-180.

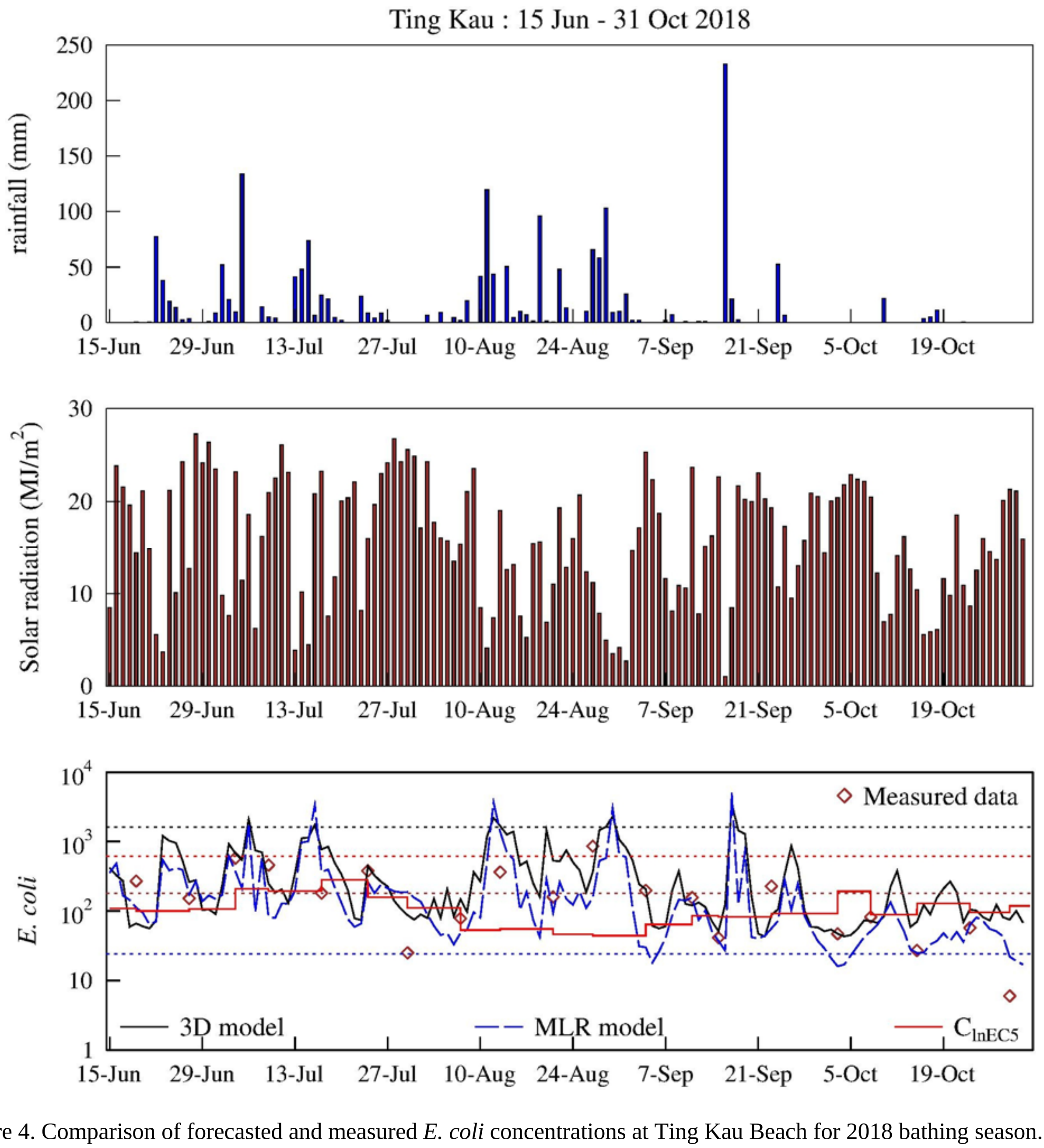

Figure 4. Comparison of forecasted and measured *E. coli* concentrations at Ting Kau Beach for 2018 bathing season.